Effect of indirect dependencies on "Maximum likelihood blind separation of two quantum states (qubits) with cylindrical-symmetry Heisenberg spin coupling"

Yannick Deville[1] and Alain Deville[2]

(1) Laboratoire d'Astrophysique de Toulouse-Tarbes, Université de Toulouse, CNRS, 14 Av. Edouard Belin, 31400 Toulouse, France. Email: ydeville@ast.obs-mip.fr
(2) IM2NP, Université de Provence, Centre de Saint-Jérôme, 13397 Marseille Cedex 20, France. Email: alain.deville@univ-provence.fr

Abstract. In a previous paper [1], we investigated the Blind Source Separation (BSS) problem, for the nonlinear mixing model that we introduced in that paper. We proposed to solve this problem by using a maximum likelihood (ML) approach. When applying the ML approach to BSS problems, one usually determines the analytical expressions of the derivatives of the log-likelihood with respect to the parameters of the considered mixing model. In the literature, these calculations were mainly considered for linear mixtures up to now. They are more complex for nonlinear mixtures, due to dependencies between the considered quantities. Moreover, the notations commonly employed by the BSS community in such calculations may become misleading when using them for nonlinear mixtures, due to the above-mentioned dependencies. In this document, we therefore explain this phenomenon, by showing the effect of indirect dependencies on the application of the ML approach to the mixing model considered in [1]. This yields the explicit expression of the complete derivative of the log-likelihood associated to that mixing model.

Keywords. Maximum likelihood estimation, blind signal separation, quantum source separation, nonlinear mixtures, indirect dependency, total derivative, partial derivative, gradient.

1 Data model

In a previous paper [1], we investigated a Blind Quantum Source (or Signal) Separation (BQSS) problem. More precisely, we aimed at restoring two quantum states (qubits) after they have been coupled, i.e. after they have been "mixed", using the classical Blind Source Separation (BSS) terminology. The considered coupling was based on a cylindrical-symmetry Heisenberg model.

We showed that repeated initializations (i.e. preparations) and measurements performed with these coupled qubits resulted in an "observation vector" which may be denoted $x = [x_1, x_2, x_3]^T$ using standard BSS notations, where ^{T} stands for transpose. In the considered problem, the components of this vector are equal to $x_1 = p_1, x_2 = p_2$ and $x_3 = p_3$, where the quantities p_j are defined in [1].

Moreover, we proved that the components of the above observation vector may be expressed as nonlinear combinations (i.e. nonlinear "mixtures", in BSS terms) of a set of "source signals". Using standard BSS notations, the vector composed of these source signals reads $s = [s_1, s_2, s_3]^T$. In the considered problem, the components of this vector are equal to $s_1 = r_1, s_2 = r_2$ and $s_3 = \Delta_I$, where the right-hand terms of these equations are defined in [1].

The "mixing model" then consists of the equations which define how the components of the observation vector are expressed with respect to (i) the components of the source vector and (ii) the mixing parameter(s). By modelling the considered quantum configuration, we showed in [1] that this configuration involves a single mixing parameter, denoted v. The mixing equations then read (see (18), (24), (25) in [1])

$$r_1^2 r_2^2 \;=\; p_1 \tag{1}$$
$$(1 - r_1^2)(1 - r_2^2) \;=\; p_2 \tag{2}$$
$$r_1^2(1 - r_2^2)(1 - v^2) + (1 - r_1^2)r_2^2 v^2$$
$$-2r_1 r_2 \sqrt{1 - r_1^2}\sqrt{1 - r_2^2}\sqrt{1 - v^2}\, v \sin \Delta_I \;=\; p_3. \tag{3}$$

This mixing model may also be expressed in compact form as

$$x = g(s) \tag{4}$$

where the nonlinear mixing function g has three components g_1 to g_3, with $x_i = g_i(s)$, $\forall i \in \{1 \ldots 3\}$. These components g_i are respectively defined by (1), (2) and (3).

Eq. (4) focuses on the signals (i.e. sources and observations). It hides the fact that the observations also depend on the parameters of the mixing model, i.e. on v in the model considered here. This additional dependency can be made explicit, by rewriting (4) as

$$x = g(s, v). \tag{5}$$

The latter form of the mixing model is better suited to the maximum likelihood approach considered below in this document.

2 Previously reported results for maximum likelihood approach

In our BSS problem, we aim at retrieving a sequence of unknown source vectors s from the corresponding sequence of measured observation vectors x and from the mixing parameter v, which is also initially unknown. This parameter v should therefore be estimated before proceeding to the source restoration step.

In [1], we investigated the estimation of v by means of the maximum likelihood (ML) approach. While we detailed this procedure in [1], we here only summarize its features which are of importance for developing new aspects of this ML approach further in this document.

The function used to estimate v is the (normalized) log-likelihood of the considered data. Under some assumptions [1], the log-likelihood here reads (see (34) in [1])

$$\mathcal{L} = \sum_{i=1}^{3} E_t[\ln f_{S_i}(s_i(t))] - E_t[\ln |J_g(s(t))|] \tag{6}$$

2

where $E_t[.]$ represents temporal averaging over the sequence of available data, $f_{S_i}(.)$ are the probability density functions of the source signals and $J_g(s)$ is the Jacobian of the mixing function g. For the function g considered in this investigation, we have (see (28) in [1])

$$J_g(s) = 8r_1^2 r_2^2 (r_2^2 - r_1^2)\sqrt{1 - r_1^2}\sqrt{1 - r_2^2}\sqrt{1 - v^2}v\cos\Delta_I. \tag{7}$$

When applying the ML approach to a parameter estimation problem, the value selected for the set of parameters to be estimated is the value which maximizes the log-likelihood $\mathcal{L}$. In order to determine the location of this maximum, one usually considers the gradient of $\mathcal{L}$ (see e.g. [2]). In our configuration, only a single parameter is to be estimated, namely v. Therefore, the gradient of $\mathcal{L}$ is here restricted to the derivative of $\mathcal{L}$ with respect to v. In [1], we denoted this gradient or derivative by using the notation most often employed in the BSS community (see e.g. [2]), i.e. $\frac{\partial \mathcal{L}}{\partial v}$. We keep this notation in this section, in order to clearly refer to the equations available in [1], but in Section 3 we will show that it may be misleading and we will therefore introduce another notation in Section 3. In [1], we showed that this gradient reads (see (37) in [1])

$$\frac{\partial \mathcal{L}}{\partial v} = -\sum_{i=1}^{3} E_t[\psi_i(s_i)\frac{\partial s_i}{\partial v}] - E_t[\frac{1}{J_g}\frac{\partial J_g}{\partial v}] \tag{8}$$

where (see (36) in [1])

$$\psi_i(u) = -\frac{\partial \ln f_{S_i}(u)}{\partial u} \quad \forall i \in \{1\ldots 3\} \tag{9}$$

are the score functions of the source signals.

The last stage of this investigation consists in deriving the expressions of all the terms of the right-hand part of (8). In [1], we showed that the terms $E_t[\psi_i(s_i)\frac{\partial s_i}{\partial v}]$ corresponding to $i = 1$ and $i = 2$ are equal to zero. The term corresponding to $i = 3$ was derived from equations (38) to (40) in [1], which read

$$\frac{\partial F}{\partial s_3}\frac{\partial s_3}{\partial v} + \frac{\partial F}{\partial v} = 0 \tag{10}$$

with

$$\frac{\partial F}{\partial s_3} = -2r_1 r_2\sqrt{1 - r_1^2}\sqrt{1 - r_2^2}\sqrt{1 - v^2}v\cos\Delta_I \tag{11}$$

and

$$\frac{\partial F}{\partial v} = 2v(r_2^2 - r_1^2) - 2r_1 r_2\sqrt{1 - r_1^2}\sqrt{1 - r_2^2}\frac{1 - 2v^2}{\sqrt{1 - v^2}}\sin\Delta_I. \tag{12}$$

For the sake of clarity, we here provide the resulting expression of $\frac{\partial s_3}{\partial v}$, which is also used below in the current document. The above equations yield

$$\frac{\partial s_3}{\partial v} = -\left[\frac{\partial F}{\partial s_3}\right]^{-1}\frac{\partial F}{\partial v} \tag{13}$$

$$= \frac{(r_2^2 - r_1^2)}{r_1 r_2\sqrt{1 - r_1^2}\sqrt{1 - r_2^2}\sqrt{1 - v^2}\cos\Delta_I} - \frac{(1 - 2v^2)\sin\Delta_I}{(1 - v^2)v\cos\Delta_I}. \tag{14}$$

The last term that should be determined to obtain the complete expression in (8), i.e. in (37) of [1], is its term $\frac{\partial J_g}{\partial v}$. In Equation (41) of [1], we provided an explicit expression

that we denoted $\frac{\partial J_g}{\partial v}$. We here aim at warning the reader that, although each of the equations (37) and (41) of [1] is correct in itself if interpreted correctly, there may be a misunderstanding when considering these equations together, because the notation $\frac{\partial J_g}{\partial v}$ does not have the same meaning in both of them. Briefly, $\frac{\partial J_g}{\partial v}$ represents a total derivative in (37) of [1], but a partial derivative in (41) of [1], as detailed below in Section 3. The right-hand expression in Equation (41) of [1] is therefore *only one of the terms* which compose the complete expression of $\frac{\partial J_g}{\partial v}$ to be used in (37) of [1]. In the following section of the current document, we clarify this point and we determine the complete expression of $\frac{\partial J_g}{\partial v}$ of (37) of [1], i.e. of (8) of the current document.

3 New results for maximum likelihood approach

When applying the ML approach to any BSS configuration, the log-likelihood $\mathcal{L}$ is considered for the fixed set of observed vectors. The only independent variable in this approach is the set of mixing parameters to be estimated, which is here restricted to v. The source vectors are dependent variables, here linked to the observations and to v by (5). The overall variations of the log-likelihood $\mathcal{L}$ with respect to v result from two types of terms contained in the expression of $\mathcal{L}$, i.e. (i) the terms involving v itself and (ii) the terms involving the source signals s_1, s_2 and s_3, which are here considered as functions of v and may therefore be denoted as $s_1(v), s_2(v)$ and $s_3(v)$ for the sake of clarity.

This approach should be kept in mind when interpreting all equations in [1], which were partly gathered in Section 2 of the current document. Especially, the log-likelihood itself, which appears in the left-hand term of (6), may be denoted as $\mathcal{L}(v, s_1(v), s_2(v), s_3(v))$ for the sake of clarity. In order to determine the location of the maximum of this log-likelihood, one should then consider the *total* derivative of $\mathcal{L}(v, s_1(v), s_2(v), s_3(v))$ with respect to v. The notations with partial derivatives in (8) may therefore be misleading, as confirmed below. Therefore, (8) should preferably be rewritten as

$$\frac{d\mathcal{L}}{dv} = -\sum_{i=1}^{3} E_t[\psi_i(s_i)\frac{ds_i}{dv}] - E_t[\frac{1}{J_g}\frac{dJ_g}{dv}] \tag{15}$$

with

$$\psi_i(u) = -\frac{d\ln f_{S_i}(u)}{du} \quad \forall i \in \{1\dots3\}. \tag{16}$$

The term $\frac{dJ_g}{dv}$ in (15) then deserves some care because, as shown by (7), the Jacobian J_g contains the above-defined two types of dependencies with respect to v, i.e. (i) *direct dependencies* due to the factors in (7) which explicitly contain v and (ii) *indirect dependencies* due to the factors in (7) which depend on the source signals, which themselves depend[1] on v in the ML approach. We here have to consider the *total* derivative $\frac{dJ_g}{dv}$,

---

[1]When applying the ML approach to a general mixing model, all source signals depend on all mixing parameters . For the specific mixing model considered here, we will show below that *only one* of the three source signals actually depends on the mixing parameter.

4

which takes into account both types of dependencies, and which therefore reads

$$\frac{dJ_g}{dv} = \frac{\partial J_g}{\partial v} + \sum_{i=1}^{3} \frac{\partial J_g}{\partial s_i} \frac{ds_i}{dv}. \tag{17}$$

In this expression, $\dfrac{\partial J_g}{\partial v}$ is the *partial* derivative of J_g with respect to v, calculated by considering that the source signals are constant. This partial derivative is the quantity that we provided in Equation (41) of [1], where we also denoted it as $\dfrac{\partial J_g}{\partial v}$. However, since we independently used the same notation $\dfrac{\partial J_g}{\partial v}$ in the last term of (37) of [1] (that we repeat as (8) of the current document), we may have incorrectly suggested to the reader that this last term of (37) of [1] is equal to the above-mentioned partial derivative. Instead, let us insist again that the partial derivative $\dfrac{\partial J_g}{\partial v}$ in (41) of [1] is first to be added with the other terms in the right-hand part of (17), in order to obtain the overall total derivative $\dfrac{dJ_g}{dv}$ defined by (17). What should eventually be used in the last term of (8) or (15) is this *total* derivative.

So, back to the calculation of all terms of the total derivative $\dfrac{dJ_g}{dv}$ in (17), the first term is available from Equation (41) of [1] and reads

$$\frac{\partial J_g}{\partial v} = 8r_1^2 r_2^2 (r_2^2 - r_1^2)\sqrt{1 - r_1^2}\sqrt{1 - r_2^2}\frac{1 - 2v^2}{\sqrt{1 - v^2}}\cos \Delta_I. \tag{18}$$

The other three terms of (17) involve the derivatives[2] $\frac{ds_i}{dv}$. Their calculation first require us to determine the expressions of the source signals $s_1 = r_1, s_2 = r_2$ and $s_3 = \Delta_I$ with respect to the observations and mixing parameter v. The first two source signals, i.e. $s_1 = r_1$ and $s_2 = r_2$ are obtained by solving (1) and (2). These equations are independent of v. Therefore, s_1 and s_2 are also independent of v and yield

$$\frac{ds_1}{dv} = 0 \quad \text{and} \quad \frac{ds_2}{dv} = 0. \tag{19}$$

The derivative $\frac{ds_3}{dv}$ was already provided above in[3] (14). That derivative is used in the term of (17) related to the third source signal, together with the partial derivative $\dfrac{\partial J_g}{\partial s_3}$, which is obtained from (7) and reads

$$\frac{\partial J_g}{\partial s_3} = -8r_1^2 r_2^2 (r_2^2 - r_1^2)\sqrt{1 - r_1^2}\sqrt{1 - r_2^2}\sqrt{1 - v^2}v \sin \Delta_I. \tag{20}$$

Inserting (18), (19), (20) and (14) in (17) yields

$$\frac{dJ_g}{dv} = 8r_1^2 r_2^2 (r_2^2 - r_1^2)\sqrt{1 - r_1^2}\sqrt{1 - r_2^2}\frac{1 - 2v^2}{\sqrt{1 - v^2}}\frac{1}{\cos \Delta_I} - 8r_1 r_2 (r_2^2 - r_1^2)^2 v\frac{\sin \Delta_I}{\cos \Delta_I}. \tag{21}$$

---

[2]Unlike $\frac{\partial J_g}{\partial v}$ considered above, the derivatives $\frac{ds_i}{dv}$ do not yield any risk of ambiguity between partial and total derivatives: each considered signal s_i is here considered independently from the other source signals and only involves a direct dependency with respect to v.

[3]Due to the type of notations employed throughout Section 2, we used a partial derivative notation in (14). Anyway, the quantity considered in that equation (14) is the same as $\frac{ds_3}{dv}$ addressed here, as explained in the previous footnote.

This completes the comment that we aimed at providing in this document, concerning the total derivative to be used in the last term of (15).

For the sake of clarity, we now conclude by providing the explicit expression of the derivative of the log-likelihood which results from the complete expression (21). Using this expression together with (7), (14) and (19) allows us to rewrite (15) as

$$\begin{aligned}
\frac{d\mathcal{L}}{dv} &= -E_t\left[\psi_{\Delta_I}(\Delta_I)\left\{\frac{(r_2^2 - r_1^2)}{r_1 r_2\sqrt{1-r_1^2}\sqrt{1-r_2^2}\sqrt{1-v^2}\cos\Delta_I} - \frac{(1-2v^2)\sin\Delta_I}{(1-v^2)v\cos\Delta_I}\right\}\right] \\
&\quad -E_t\left[\frac{(1-2v^2)}{(1-v^2)v\cos^2\Delta_I} - \frac{(r_2^2-r_1^2)\sin\Delta_I}{r_1 r_2\sqrt{1-r_1^2}\sqrt{1-r_2^2}\sqrt{1-v^2}\cos^2\Delta_I}\right].
\end{aligned} \quad (22)$$

References

[1] Y. Deville, A. Deville, "Maximum likelihood blind separation of two quantum states (qubits) with cylindrical-symmetry Heisenberg spin coupling", Proceedings of the 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008), pp. 3497-3500, Las Vegas, Nevada, USA, March 30 - April 4, 2008.

[2] A. Hyvärinen, J. Karhunen, E. Oja, "Independent Component Analysis", Wiley, New York, 2001.